

This is a repository copy of *Instructor personality matters for student evaluations: : Evidence from two subject areas at university*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/126000/>

Version: Accepted Version

Article:

Kim, Lisa orcid.org/0000-0001-9724-2396 and MacCann, Carolyn (2017) Instructor personality matters for student evaluations: : Evidence from two subject areas at university. *British Journal of Educational Psychology*. pp. 1-22. ISSN 0007-0998

<https://doi.org/10.1111/bjep.12205>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Abstract

Background. Instructors are under pressure to produce excellent outcomes in students. Although the contribution of *student* personality on student outcomes is well established, the contribution of *instructor* personality to student outcomes is largely unknown.

Aim. The current study examines the influence of instructor personality (as reported by both students and instructors themselves) on student educational outcomes at university.

Sample and method. Mathematics and psychology university students ($N = 515$) and their instructors ($n = 45$) reported their personality under the Big Five framework.

Results. Multilevel regressions were conducted to predict each outcome from instructor personality, taking into account the effects of student gender, age, cognitive ability, and personality, as well as instructor gender and age. Student-reports of instructor personality predicted student evaluations of teaching but not performance self-efficacy or academic achievement. Instructor self-reports did not predict any of the outcomes. Stronger associations between student-reports and the outcomes than instructor self-reports could be explained by students providing information on both the predictor and the outcome variables, as well as a greater number of raters providing information on instructor personality. Different domains of the instructor Big Five were important for different element of student evaluations of teaching.

Conclusions. The study highlights the importance of studying instructor personality, especially through other-reports, to understand students' educational experiences. This has implications for how tertiary institutions should use and interpret student evaluations.

KEYWORDS: instructor non-cognitive characteristics; instructor personality; Big Five; five-factor model of personality; student evaluations of teaching

DATE OF SUBMISSION: 21 November 2016

DATE OF RESUBMISSION: 4 August 2017, 11 October 2017

Instructor Personality Matters for Student Evaluations: Evidence from Two Subject Areas at University

Instructors critically shape student educational experiences and outcomes (Aaronson, Barrow, & Sander, 2007; Kane, Rockoff, & Staiger, 2008). Although consistent effects of student personality (especially conscientiousness) on student academic outcomes in both K–12 and in university (Poropat, 2009) are reported, there is limited research on instructor personality (Klassen & Tze, 2014). We examine instructor personality an under-studied characteristic that can be useful for understanding university students' educational outcomes.

Non-cognitive characteristics are increasingly recognized as important qualities in educational research (Heckman & Kautz, 2012; Richardson et al., 2012). Personality is one set of these characteristics, which underlies an individual's cognition, emotions, and behavior that distinguishes one individual from another (John & Srivastava, 1999). The Big Five is the dominant personality model, which proposes five domains underlying individuals' personality: openness (imaginative, curious, cultured), conscientiousness (orderly, responsible, dependable), extraversion (talkative, assertive, energetic), agreeableness (good-natured, cooperative, trustful), and emotional stability (calm, secure, unemotional; John & Srivastava, 1999).

Despite the long history of interest in the personality profile of effective instructors (e.g., Dodge, 1943), there is a lack of systematic investigations into instructor personality using established frameworks. The growing consensus that personality research is relevant to educational outcomes (Heckman & Kautz, 2012) now warrants thorough applications in another group of individuals in the educational field — the instructors.

Instructor Personality and Educational Outcomes

Models of teaching and learning recognize that teaching is an interpersonal process in which instructor characteristics manifest in the classroom to influence students' classroom

outcomes (Murray et al., 1990). Indeed, Hattie's (2009) review indicates that 'teacher factors' have the greatest impact on student learning and achievement compared to home, curriculum, student, or school factors. Dunkin and Biddle's (1974) model of classroom teaching posits four classes of variables: (1) presage variables (instructor experiences, training, and characteristics); (2) context variables (student experiences and characteristics); (3) process variables (what instructors and students do in the classroom); and (4) product variables (immediate and long term educational outcomes). Both presage variables and context variables impact the process variables, which in turn influence the product variables. Similarly, Groccia's (2012) model of teaching and learning proposes that instructor variables, learner variables, learning process and context, and course content affect instructional processes, which in turn affect learning outcomes. We believe that instructor personality constitutes a key presage variable and instructor variable in Dunkin and Biddle's (1974) model and in Groccia's (2012) model, respectively, which affects student outcomes.

For example, an instructor who has high levels of agreeableness will generally show understanding and warmth across different classroom situations, including when a student reacts emotionally to an exam result and when students are in heated conflict. Such instructor behavior would affect how students perceive the instructor and the subject, and possibly how well believe they will do in the subject and how well they actually do in it. Thus, both models of learning and personality provide a theoretical rationale for our hypothesis that instructor personality influences student educational outcomes.

Measures of Student Educational Outcomes

The few prior studies of instructor personality have tended to focus on a single criterion (e.g., Patrick, 2011; Rockoff, Jacob, Kane, & Staiger, 2011) rather than a broad criterion space of multiple inter-related outcomes. However, instructor personality domains may affect the outcomes differently. Accordingly, we examine the influence of instructor

personality on three classes of outcomes: (a) student evaluations of teaching, (b) performance self-efficacy, and (c) academic achievement. The three classes of outcomes were chosen as they target different aspects of student educational outcomes. Namely, student evaluations of teaching reflect the students' perceptions of the teacher, performance self-efficacy reflects the students' perceptions of themselves, and academic achievement is external to the students' perceptions.

Student evaluations of teaching. Measures of instructional quality have received international attention (e.g., OECD, 2013) given their important links with learning outcomes (Hattie, 2009). At university, student reports of instructional quality (i.e., student evaluations of teaching) have increasingly been used as measures of university instructor performance (Marsh & Roche, 1997) and numerous studies have been published on the psychometric properties and usefulness student evaluation of teaching tools, such as the Student Evaluations of Educational Quality (e.g., Marsh, 1984; Marsh, 2007; Marsh & Overall, 1980). Student evaluations have then been used for the purpose of marketing courses (Richardson, 2005), providing feedback to instructors (Taylor & Tyler, 2012), and even as part of making high-stakes administrative decisions (Murray et al., 1990). Given the increased international use of student evaluations in higher education, it is important to examine the extent to which these evaluations are impacted by instructor non-cognitive characteristics.

The Student Evaluation of Educational Quality (SEEQ; Marsh, 1982) is one of the most commonly-used tools assessing multidimensional elements of student evaluations of teaching in published work (Richardson, 2005). The measure assesses 11 elements of teaching, seven of which particularly pertain to instructors¹. They describe the level of

¹ The other four measures (breadth, workload/difficulty, examinations, and assignments) were excluded from the current study as the syllabus breadth was fixed across teachers and students did not receive their exams or assignments at the time the data were collected.

learning experienced during the course (learning focus), instructor enthusiasm in the classroom (enthusiasm), clarity and organization of classroom material (organization), encouragement of discussion and participation in the classroom (group interaction), instructor accessibility and friendliness to students (individual rapport), overall course evaluation compared to other courses (overall course evaluation), and overall instructor evaluation compared to other instructors (overall instructor evaluation). SEEQ's statistical properties, including dimensionality, reliability, and validity have been shown to be strong in many studies (see for a summary, Marsh, 2007).

Very few studies have investigated the link between instructor personality and multiple elements of student evaluations, although existing studies support its link. Qualitative studies indicate that university faculty often believe that student evaluations are strongly associated with instructor non-cognitive qualities such as personality or likability (Simpson & Siguaw, 2000). Quantitative studies also suggest that different personality domains are important in predicting different elements of student evaluations. In relation to individual rapport and group interaction, Isaacson et al. (1963) found that university teaching fellows rated as high in surgency (similar to extraversion) received high ratings. Rapport with students was also highly correlated with university instructors' personal warmth (Murray, 1975), which is an element of agreeableness. In relation to overall instructor evaluations, Patrick (2011) found that conscientiousness was the strongest predictor, even after controlling for students' previous learning and expected grade. Similarly, Murray (1975) reported that leadership and objectivity (both elements of conscientiousness) were the strongest correlates of overall instructor rating. In relation to overall course evaluations, Patrick (2011) reported that openness was the strongest predictor of overall class ratings, even after controlling for students' previous learning and expected grade.

In addition to these previous findings, we expect that instructors with high levels of certain personality domains will demonstrate behaviors, which are associated with receiving high student evaluations on certain elements. Specifically, extraverted instructors are likely to receive high evaluations on enthusiasm in the class, and conscientious instructors are likely to receive high evaluations on organization of the class material. In this way, we can make predictions about how instructor personality domains may be associated with various elements of student evaluations of teaching that is assessed using a well-validated multi-component student evaluation instrument.

Performance self-efficacy. Performance self-efficacy (PSE) is a type of self-efficacy capturing one's belief about their capability to perform academically. A meta-analysis reported that of a range of non-cognitive competencies, PSE showed the strongest correlation with grade point average (Richardson et al., 2012). Despite the importance of the construct, its association with instructor personality is yet unknown. Among the Big Five, neuroticism (the opposite of emotional stability) captures one's disposition to experience negative emotions, such as anxiety and self-consciousness. In the educational context, instructors may verbalize their anxious thoughts and may even verbalize or show their lack of confidence in their teaching skills or in their students mastering the material. These behaviors and words can negatively affect students—they may come to believe the instructors' thoughts, model similar qualities to the instructor, and focus on negative thoughts and emotions. These may culminate to students having low PSE. This hypothesis is bolstered by meta-analytic evidence that low emotional stability is associated with poor job performance (Barrick & Mount, 1991; Salgado, 1997). Thus, we expect that low instructor emotional stability may be associated with low student PSE.

Academic achievement. Academic achievement, measured mostly as grades and grade point averages, is one of the most common proxies for student learning. Although

studies clearly indicate that *student* personality predicts academic achievement (Poropat, 2009), it is unclear whether *instructor* personality also predicts academic achievement. As student learning can be considered a measure of instructor job performance, examining meta-analyses on personality and job performance can be useful. Studies indicate that conscientiousness is the strongest personality predictor of job performance (Barrick & Mount, 1991; Judge, Rodell, Klinger, Simon, & Crawford, 2013; Salgado, 1997).

Two studies have explicitly examined the link between instructor personality and student academic achievement. Garcia, Kupczynski, and Holland (2011) found that English language arts instructors with high levels of self-reported conscientiousness had tenth grade students with higher Texas Assessment of Knowledge Skills scores. On the other hand, Rockoff et al., (2011) found that instructor self-reported extraversion and conscientiousness did not predict fourth to eighth grade student math scores when student and school characteristics were controlled. Accordingly, we thus test whether high instructor conscientiousness is associated with student academic achievement, operationalized as the final course mark.

Sources of Personality Report

Instructor personality can be reported by the instructors themselves or by others, such as students and colleagues. Clear pragmatic and conceptual differences distinguish self-reports from other-reports of personality. Self-reports are the most common source of personality measurement partly due to their convenient methodology. Their common usage is also partly due to the assumption that the self is the best informant — the self has access to a unique perspective of their own private experiences, such as physiological changes, cognitions, and history of behaviors across different environments (Funder, 1999; John & Robins, 1993). However, defensiveness and self-presentational strategies may cause inaccuracies in personality self-reporting that may undermine the construct validity of self-

reports (Morgeson et al., 2007; Paulhus & Vazire, 2007). Other-reports are an alternative source of personality report and are assumed to be less susceptible to distortions and less affectively charged than self-reports (Kolar, Funder, & Colvin, 1996) and are hence more internally consistent (Balsis, Cooper, Oltmanns, 2015). Moreover, meta-analyses indicate that other-reports of personality show a much stronger prediction of academic achievement than self-reports. Poropat's (2014) meta-analysis found that correlations with academic achievement ranged from .05 to .38 for other-reports and -.02 to .22 for self-reports. Connelly and Ones' (2010) meta-analysis found that such correlations ranged from .01 to .41 for other-reports and -.02 to .25 for self-reports. The overall effect size for predicting academic achievement from other-reported personality is on par with the largest effect sizes of any predictor reported in meta-analyses (c.f., Hattie, 2009). Similarities and differences between self-ratings and other-ratings can be understood within the Trait-Reputation-Identify Model (McAbee & Connelly, 2016). This model proposes that the agreement between self- and other-ratings represents personality *traits*, the unique perspective of other-raters represents the target's personality *reputation*, and the unique perspective the individual holds of themselves represents their personality *identity*.

Although very few studies have compared different sources of teacher personality reports, they indicate that other-reports have stronger effects than self-reports. For example, Isaacson et al. (1963) reviewed university teaching fellows' teaching ability as reported by their students and assessed its association with personality traits as rated by other teaching fellows and as rated by the self. Of the five personality traits reported by other teaching fellows, those who received high ratings on culture (similar to openness) and agreeableness received higher ratings of teaching ability. In contrast, of the self-reported 21 personality traits, those who reported high levels of enthusiasm received higher ratings of teaching ability and the effect size was slightly smaller compared to other-reports. In a similar study,

Feldman (1986) reviewed college teachers' teaching effectiveness as reported by their students. When teacher personality was self-reported, four of the 14 clusters of personality traits were significantly correlated with overall student evaluation and they were of small effect sizes. When teacher personality was reported by others (colleagues and students), 11 trait clusters were significantly correlated with overall student evaluations and they were of moderate to large effect sizes. These two studies indicate that student-reported teacher personality may be more strongly associated with the outcomes than self-reported personality.

Other-reports may be more strongly associated with the outcomes for a couple of reasons. First, when the same individual provides information on both the predictor and the outcome, the predictor–outcome correlation is higher than when information on the predictor and outcome are obtained from different individuals (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). This is known as the common rater effect. If this effect holds, then student-reports of teacher personality may show stronger associations with student-reported outcomes (i.e., SEEQ and PSE), as compared to self-reports of teacher personality. We expect that when the same individual provides information on both the predictor and outcome, the predictor–outcome correlation will be greater than when this information is provided by different individuals. Second, other-reports may provide more reliable information when there are multiple other-reporters compared to a single self-reporter. We will refer to this as the multiple rater effect. Connelly and Ones (2010) noted the importance of multiple other-reporters to overcome the idiosyncrasies of single ratings and to increase reliability. To test this hypothesis, we will randomly select a single student-reporter for each instructor and compare this to mean student-reported instructor personality in terms of their correlations with the outcome variables. We expect that there will be a statistical difference, indicating that a greater number of other-raters contribute to the strength of its relationship with the outcomes.

Current Study

The current research investigates the role of university instructor personality on student evaluations of teaching, PSE, and academic achievement. We consider student-reports and self-reports of instructor personality to examine whether the source of reporting affects the relationship between teacher personality and teacher effectiveness.

Our study includes several advances over previous research. First, we consider two sources of report of instructor personality: self-report and student-report, given their potential different predictive validities. Second, we use multiple student educational outcome measures (student evaluations of teaching, student PSE, and student course mark) so as to assess how different instructor personality domains may be associated with these differently. Third, as student personality, gender, age, and cognitive ability as well as instructor gender and age can be associated with educational outcomes (Marsh, 2007; Poropat, 2009; Sabbe & Aelterman, 2007), our study examines the incremental predictive validity of instructor personality to investigate its pragmatic utility.

Summary of Research Hypotheses

Given the findings from previous studies, instructor personality is expected to significantly predict student educational outcomes, with student-reports of instructor personality showing stronger predictions than instructor self-reports (H1). We also expect that there will be evidence for the two explanations as to why student-reports of instructor personality could be stronger predictors than instructor self-reports of personality: (a) the common rater effect (H2a); and (b) the multiple rater effect (H2b).

We also posit hypotheses about the associations between instructor Big Five and the outcomes. For student evaluations, we hypothesize that: (a) teacher openness predicts student learning focus, and overall course evaluation (H3a); (b) instructor conscientiousness predicts instructor organization, and overall instructor evaluation (H3b); (c) instructor extraversion

predicts instructor enthusiasm, and group interaction (H3c); and (d) instructor agreeableness predicts individual rapport (H3d). For PSE, we hypothesize that instructor emotional stability is a positive predictor (H3e). For academic achievement, we hypothesize that instructor conscientiousness is a positive predictor (H3f).

Method

Participants

The participants were tutors (instructors teaching tutorial classes) and students from first year mathematics and psychology tutorial classes (classes aimed at reviewing and expanding on lecture material). Tutors must hold an undergraduate degree, and are either academic staff or post-graduate students currently completing a PhD in the content area. The ages of the 515 students included in the study ranged from 17 to 55 ($M = 19.31$, $SD = 3.19$; 69.71% female). Student reported information on 56 mathematics across 97 classes and 32 psychology tutors across 103 classes. This was more than the number of instructors who provided self-reports: 27 mathematics instructors and 18 psychology instructors, whose ages ranged from 21 to 68 ($M = 29.24$, $SD = 11.51$; 46.67% female).

Test Battery

Self-rated instructor personality. Instructors reported their own personality using Saucier's (1994) 40-item shortened version of Goldberg's (1992) personality marker inventory. This inventory consists of eight adjectives assessing each of the five personality domains, namely openness, conscientiousness, extraversion, agreeableness, and emotional stability. An example of an item for conscientiousness is "Practical." Students reported how accurately each adjective described their personality by assigning a number from 1 (*Extremely Inaccurate*) to 9 (*Extremely Accurate*).

Course mark. Students' end of semester course mark in the mathematics or psychology unit for which they rated their tutor was collected from official university records

after semester ended. Course marks can range from 0 to 100 within which different grading can be given: Fail (0-49), Pass (50-64), Credit (65-74), Distinction (75-84), and High Distinction (85 and over). Of the students who pass, the largest proportion will be awarded a Pass grade, followed by a Credit, followed with a Distinction, with High Distinctions rarely awarded to more than 5% of the unit of study.

Students completed the measures below.

Analogies test. Cognitive ability was assessed using 15 analogical reasoning items from MacCann, Joseph, Newman, and Roberts (2014). Students were shown a word pair representing a particular relationship then asked to select which of the five word-pairs demonstrated the same relationship as the initial set. An example is “SEDATIVE : DROWSINESS” and students need to select one of the following options: “(a) epidemic : contagiousness; (b) vaccine : virus; (c) laxative : drug; (d) anesthetic : numbness; (e) therapy : psychosis”.

Self-rated student personality. Students reported their own personality using Saucier’s (1994) 40-item shortened version of Goldberg’s (1992) personality marker inventory, as described above.

Student-rated instructor personality. Students reported their instructors’ personality using the same adjective-based scale used to rate their own personality (Saucier, 1994). Instructions were adapted to refer to the relevant instructor, specifically “Think of your CURRENT FIRST YEAR MATHEMATICS/PSYCHOLOGY TUTOR. Use this list of common human traits to describe your CURRENT FIRST YEAR MATHEMATICS/PSYCHOLOGY TUTOR as accurately as possible.” A manipulation check question after each personality questionnaire asked participants to recall the target of the personality questionnaire. The order of the personality questionnaires was counter-balanced.

Student evaluations of educational quality (SEEQ). The 31-item instrument measures student evaluations of teaching (Marsh, 1982). The seven subscales included in the study were: learning focus (4-items), enthusiasm (4-items), organization (2-items), group interaction (4-items), individual rapport (4-items), overall evaluation of the course (1-item), and overall evaluation of the instructor (1-item). An example of an item for organization is “The tutor’s explanations were clear”. To each item, students responded on a scale from 1 (*Very Poor*) to 5 (*Very Good*).

PSE. Participants reported the mark out of 100 that they expected to receive as their final subject course mark.

Procedure

Mathematics students across eight first year mathematics courses were emailed an invitation to participate in the survey in exchange for the chance to win one of ten movie tickets. Mathematics instructors were e-mailed an invitation to participate in the survey with coffee vouchers as incentives. Psychology students in a first year psychology course participated in the study as part of the five hours of research that they must participate in from a registered database of multiple studies for course credit. Psychology instructors were e-mailed an invitation to participate in the survey with no incentives for their participation.

Unproctored online surveys were available to students and instructors from week 9 of a 13-week semester. 154 mathematics students and 443 psychology students completed the survey. However, 17 mathematics students and 65 psychology students were excluded as they demonstrated a non-variant responding pattern in one or more of the measures, took less than half the designated time of 30 minutes, scored less than a fifth of the items correct on the analogies test, and/or failed the manipulation check questions. Course marks were available for 134 mathematic students and 322 psychology students. All responses from the instructors

were included. Both protocols were approved by the Human Research Ethics Committee at the last author's institution.

Statistical Analyses

Similarities between two sources of instructor personality report were examined firstly by calculated their mean differences using Hedge's g for 45 instructors. The statistical significance of these differences was calculated using paired samples t -tests. The similarities between the two sources of instructor personality report were also examined using Pearson correlations.

Two hypotheses as to why student-reports may provide higher correlations than instructor self-reports were tested using Steiger's z -test for dependent correlations (Steiger, 1980). First, we tested the common rater effect using the random assignment design. Within each instructor, students were randomly assigned to provide information on either the subjective outcomes (i.e., SEEQ and PSE) or the instructor's personality. That is, predictor and outcome variables were based on information from different sets of students. Second, we tested the multiple rater effect for each of the 88 instructors students provided an instructor personality rating. We randomly selected a single student-report of their instructor personality and compared it to the mean student-reports of the instructor personality.

Intra-class correlations (ICCs) indicate the proportion of variance at each level in the null model for each outcome. The ICCs were .17 for learning focus, .17 for enthusiasm, .12 for organization, .24 for group interaction, .08 for individual rapport, .07 for overall evaluation of the course, .16 for overall evaluation of the instructor, .05 for PSE, and .05 for course mark. Although some of the values were smaller than the .10 guideline suggested for educational research (Hox, 2002), modeling the nested data in a non-hierarchical manner could produce biased estimates. Accordingly, multilevel regressions were conducted to predict SEEQ subscales, PSE, and course mark. As has been noted in other educational

studies (Jacob, Zhu, & Bloom, 2010), very low ICCs can produce high R^2 (often 1.00) in higher levels (in our case level 2). At level 1 of the multilevel regressions, student gender and age were entered as well as cognitive ability and student Big Five (with grand mean centering). At level 2, instructor gender and age were entered as well as one instructor Big Five domain (with grand mean centering). We are interested in the overall differences of student personality (and student ratings of instructor personality) relative to the whole student sample, not the relative effect of student personality (or student ratings of instructor personality) within each instructor. For this reason, we used grand mean centering of level 1 variables rather than group mean centering.

Results

Reliability and Descriptive Statistics

Table 1 shows the reliability and descriptive statistics for student personality (self-reported) and instructor personality (both student-reported and self-reported). For student-reports of instructor personality, the mean of the student-reports for each of the 88 instructors was calculated to represent the level 2 values for student-reported instructor personality (in line with the multilevel regressions reported later). The scale reliability estimates are reported in Cronbach alphas and inter-rater reliabilities for individual student-reported instructor personality are reported in ICCs. Reliability estimates ranged from .78 to .85 for student self-reported personality, .79 to .89 for student-reported teacher personality, and .79 to .90 for teacher self-reported personality. In addition, the mean analogies score was 9.42 ($SD = 2.90$), and its Cronbach's alpha reliability was .69.

INSERT TABLE 1 ABOUT HERE

Differences between Two Sources of Instructor Personality

Mean differences between the two sources of instructor personality report were significant for conscientiousness and emotional stability (see Table 1). Specifically, the means of the student-reports were higher than instructor self-reports with moderate to large effect sizes. The two sources of instructor personality were significantly and positively associated with each other only for conscientiousness (see Table 1). Although the direction of the other correlations were positive (except extraversion), the magnitudes were much smaller than previous research (e.g., Connelly & Ones, 2010).

Correlations between Personality and Educational Outcomes

Table 2 contains the zero-order correlations of student and instructor personality with the mean outcomes for each instructor, as well as the reliability and descriptive statistics for all outcome variables. Instructors who reported themselves to be low in extraversion and emotional stability had students with higher PSE, with small-moderate effect sizes. For mean student-reported instructor personality, all five personality domains showed small to moderately large positive significant correlations with most SEEQ subscales. PSE was positively associated with mean student-reported instructor openness, and course mark was not significantly associated with any of the personality domains. While interpreting the significance of multiple correlation coefficients may be subject to Type I error, note that 53 of the 135 correlations were significant at $p < .05$ whereas only 7 would be significant if the null hypothesis were true.

INSERT TABLE 2 ABOUT HERE

We compared the strengths of the correlations between student-reports and instructor self-reports with the outcomes using Steiger's z -test for dependent correlations (Steiger,

1980). Mean student-reports of instructor Big Five showed stronger associations with the outcomes than instructor self-reports for 40 of the 45 analyses (significantly so for 35 analyses [$2.00 \leq z \leq 4.62$; $p < .05$]; non-significantly for 5 analyses [$0.85 \leq z \leq 1.63$, $p > .05$]). Instructor self-reports showed stronger associations with the outcomes than mean student-reports for 5 of the 45 analyses with the (significantly so for 3 analyses [$2.00 \leq z \leq 2.87$, $p < .05$]; non-significantly for 2 analyses [$z = 0.05$ and 1.39 , $p > .05$]). Thus, mean student-reports seem to be a stronger predictor of the outcomes than instructor self-reports, in support of H1.

The common rater effect and the multiple rater effect were tested, respectively. Table 3 shows the correlations between randomly assigned student outcomes with single student-report of instructor personality. Although the significance of each of these correlation coefficients is not related to our study hypotheses, we note that Type I error may inflate the number of significant correlations obtained. However, 66 of the 85 correlation coefficients were significant at $p < .05$ whereas only 4 of the 85 correlations would be significant if the null hypothesis were true. When comparing the original analyses to the random assignment design, all but one of the 30 predictor–criterion relationships were significantly smaller, and the effect sizes of these were large ($2.10 \leq z \leq 9.69$, $p < .05$). Though not statistically significant, emotional stability still predicted PSE in the predicted direction ($z = 1.88$, $p > .05$). Thus, the common rater effect seems to be one of the explanations for the strength behind student-reports and outcomes, in support of H2a.

INSERT TABLE 3 ABOUT HERE

Table 3 shows the correlations between means student outcomes with single student-report of instructor personality using the single student-reporter design. Mean student-reports

of instructor personality showed stronger associations with the outcomes for 34 of the 45 analyses (significantly so for 8 analyses [$-1.99 \leq z \leq -2.51, p < .05$]; non-significantly for 16 analyses [$-1.72 \leq z \leq 1.75, p > .05$]). Single student-reports of personality showed stronger associations with outcomes for the remaining 11 analyses but non-significantly ($0.12 \leq z \leq 1.01, p > .05$). Thus, the multiple rater effect seems to be another explanation for the strength behind student-reports and outcomes, in support of H2b.

Multilevel Regressions Predicting Educational Outcomes from Personality

The level 1 R^2 differed depending on which predictor was in level 2 and hence level 1 R^2 ranges are reported in Tables 4 and 5. Table 4 reports the standardized regression coefficients when predicting the outcomes from mean student-reported instructor personality and Table 5 reports the standardized regression coefficients when predicting the outcomes from self-reported instructor personality. Tables 1A and 2A in the Appendices contain the 95% Confidence Intervals for the random slopes of the predictors.

INSERT TABLE 4 ABOUT HERE

Student-reported instructor personality. Mean student-reported instructor personality significantly predicted all SEEQ subscales, except for Enthusiasm, explaining 97 to 100% of the level 2 variance. The instructor personality domains most strongly predicting each subscale were generally in line with hypotheses (H3a-d). That is, instructor conscientiousness predicted organization, agreeableness predicted individual rapport, and openness predicted overall course evaluation. Although non-significant, the strongest predictor for learning focus was openness and the strongest predictor for enthusiasm was extraversion, as expected. Contrary to expectation, instructor emotional stability was the strongest predictor of group interaction, although non-significant. Instructor

conscientiousness predicted overall instructor evaluation, as expected, but it was more strongly predicted by agreeableness.

INSERT TABLE 5 ABOUT HERE

Contrary to expectations, PSE was not predicted by emotional stability (H3e), and academic achievement was predicted by openness (H3f) though the model R^2 was not statistically significant. Some negative regression coefficients were found throughout the models (e.g., individual rapport with emotional stability, PSE with conscientiousness, academic achievement with conscientiousness). The positive or near zero correlations but negative regression coefficients indicate suppression effects, possibly due to the variance amongst the facets within a personality domain.

Self-reported instructor personality. Self-reported instructor personality did not significantly predict any of the educational outcomes, except for the negative association between SEEQ enthusiasm subscale and openness, though the model R^2 was not statistically significant.

Discussion

The current study compared two sources of instructor personality reports and their associations with university student educational outcomes. This study highlights three key findings. First, student-reports are stronger predictors of student educational outcomes than instructor self-reports. Second, the strength of student-reports seems to be due to both the common rater effect and the multiple rater effect. Third, different instructor personality domains are associated with different elements of student evaluations.

Student-reports of instructor personality predicted student evaluations, consistent with previous research (Patrick, 2011), but did not predict PSE or academic achievement. Models

of teaching and learning (e.g., Dunkin & Biddle, 1974; Groccia, 2012) indicate various factors influence student outcomes. In the case for student self-efficacy and achievement, factors other than the instructor may play important roles, such as parental and peer influences and hours spent on independent learning. On the other hand, student evaluations of teaching seem to be more directly influenced by instructor personality. This finding indicates a greater proximal distance with the predictor, compared to the other two outcomes, as well as the importance of the combination of *trait* and *reputation* rather than *identity* in prediction within the Trait-Reputation-Identity Model (2016). That is, it may be the reputation element of personality that is important for student evaluations, rather than the identity element (or arguably the trait element).

The significant outcome associations for student-reports but not self-reports of instructor personality are consistent with previous meta-analytic findings that other-reports are stronger predictors of educational outcomes than self-reports (Connelly & Ones, 2010; Poropat, 2014). The common rater effect (Podsakoff et al., 2003) and the multiple rater effect (Connelly & Ones, 2010) seem to have contributed to why student-reports were more associated with the outcomes than self-reports. Our finding indicates that it is advantageous to obtain multiple reports from students when assessing instructor personality.

As expected, different instructor personality domains predicted different elements of students' evaluations of educational quality. That is, when the students reported their instructor's personality, the strongest associations between personality and student evaluations were for openness with learning focus, extraversion with enthusiasm, conscientiousness with organization, agreeableness with individual rapport, openness with overall course evaluation, and for agreeableness with the overall instructor evaluation. Group interaction was most strongly predicted by emotional stability not extraversion contrary to expectation. However, the magnitude of the correlation between group interaction and

instructor extraversion, agreeableness, and emotional stability were very similar ($.47 \leq r \leq .50$), thus possibly indicating a suppression effect.

Student evaluations is regarded as an important management issue for institutions as they aim to retain current students and recruit new students (Helgesen & Nasset, 2007). For some countries, like Australia, education is an important source of export for the national economy (UNESCO Institute for Statistics, 2016), and institutions have, therefore, become sensitive to students' satisfaction with their learning experience (Lala & Priluck, 2011). Some researchers challenge the validity of student evaluations as they claim such measures assess likeability rather than teaching effectiveness (Clayson & Haley, 2011). Our results, together with previous findings (e.g., Costin & Grush, 1973), challenge this idea as students' report of instructor personality traits were differentially related to classroom evaluations, such that one personality domain that was strongly associated with one classroom aspect was not associated with other classroom aspects. In this light, institutions could assist university instructors to become more aware of their own personality and how their students perceive them. Such awareness may assist in modifying or capitalizing on their teaching practices and behaviors that are beneficial to them and the students.

The effect of instructor personality on student outcomes may differ depending on the level of education. In the case of student personality, Poropat (2009) noted the influence of student personality is stronger in elementary school compared to higher education (except for conscientiousness, which has a strong effect at all levels of education). One proposed explanation was that students encounter a greater number of different learning environments and activities as they progress to higher levels of education. Therefore, the influence of any one predictor decreases as the criterion becomes more complex. Indeed, Author Citation (2017) have found that student-reports of Grades 7 to 9 teacher personality predicts student perceptions of academic and personal support from the teacher but again not PSE nor

academic achievement. The effect of instructor personality may have a stronger effect in primary education. Students in the early grades tend to have a single instructor that all in-class hours revolve around, such that the effect of instructor personality may be more potent than in a university environment where students have a much more limited number of contact hours with any individual instructor. In this light, future studies should investigate the influence of instructor personality at earlier stages of education.

The current study included Australian students undertaking psychology and mathematic courses. Future studies can strengthen the generalizability of the findings in two ways. Although the current study examined two different subject areas, future studies should consider examining students undertaking a greater diverse of subject areas. Second, the influence of teacher personality on student educational outcomes may differ across countries. The dominance of a Western, educated, industrialized, rich and democratic (WEIRD) sample in human behavior and psychology studies (Henrich, Heine, Norenzayan, 2010) is also the case for teacher personality studies. Examining the effect of teacher personality across multiple countries could indicate which personality domains are commonly important or unique in predicting student educational outcomes. Given that student perceptions of the assessment constructs may differ across countries, the measurement model should be adjusted for a more valid cross-cultural comparison as Scherer, Nilsen and Jansen (2016) noted in their measurement models of student evaluations on PISA 2012 data.

Multitrait-Multimethod analysis allows separation of the effects of personality ratings, the method factor, and error components (Eid et al., 2008). However, this type of analyses is often affected by problems such as “improper solutions (e.g., negative variance estimates), non-convergent solutions, and identification problems” (Eid et al., 2008, p. 251). We experienced all three problems when attempting to conduct this type of analysis, and thus the estimates could not be trusted. Modelling a simplified version of a multilevel SEM still led to

the same problems. These problems may be associated with having a relatively small number of level 2 clusters ($n = 45$). Future studies are encouraged to conduct this type of analysis for data with a large enough number of clusters.

Our findings highlight the importance of capturing instructor personality to understand different aspects of student educational outcomes, whereby instructor personality is strongly associated with student evaluations of teaching. As more research is dedicated to understanding the effects of instructor non-cognitive characteristics, this understanding can pave the way to improve outcomes that benefit not only the students, but also the instructors and the educational system.

References

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25, 95–135.
<http://doi.org/10.1086/508733>
- Balsis, S., Cooper, L. D., & Oltmanns, T. F. (2015). Are informant reports of personality more internally consistent than self reports of personality? *Assessment*, 22, 399–404.
<http://doi.org/10.1177/1073191114556100>
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1–26.
<http://doi.org/10.1111/j.1744-6570.1991.tb00688.x>
- Clayson, D. E., & Haley, D. A. (2011). Are students telling us the truth? A critical look at the student evaluation of teaching. *Marketing Education Review*, 21, 101–112.
<http://doi.org/10.2753/mer1052-8008210201>
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136, 1092–1122. <http://doi.org/10.1037/a0021212>
- Costin, F., & Grush, J. E. (1973). Personality correlates of teacher-student behavior in the college classroom. *Journal of Educational Psychology*, 65, 35–44.
<http://doi.org/10.1037/h0034814>
- Dodge, A. F. (1943). What are the personality traits of the successful teacher? *Journal of Applied Psychology*, 27, 325–337. <http://doi.org/10.1037/h0062404>
- Dunkin, M. J., & Biddle, B. J. (1974). *The study of teaching*. New York, NY: Holt, Rinehart and Winston.
- Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for

- different types of methods. *Psychological Methods*, 13, 230-253. <http://doi.org/10.1037/a0013219>
- Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review and synthesis. *Research in Higher Education*, 24, 139–213. <http://doi.org/10.1007/bf00991885>
- Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.
- Garcia, P. L. S., Kupczynski, L., & Holland, G. (2011). The impact of teacher personality styles on academic excellence of secondary students. *National Forum of Teacher Education Journal*, 21(3), 1–8.
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4, 26–42. <http://doi.org/10.1037/1040-3590.4.1.26>
- Groccia, J. E. (2012). A model for understanding university teaching and learning. In J. E. Groccia, M. A. T. Alsudairi, & W. Buskist (Eds.), *Handbook of college and university teaching: A global perspective* (pp. 2–15). Los Angeles, CA: SAGE Publications.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19, 451–464. <http://doi.org/10.1016/j.labeco.2012.05.014>
- Henrich, J., Heine, S., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466, 29. <http://doi.org/10.1038/466029a>
- Helgesen, Ø., & Nettet, E. (2007). What accounts for students' loyalty? Some field study evidence. *International Journal of Educational Management*, 21, 126–143. <http://doi.org/10.1108/09513540710729926>
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence

Erlbaum Associates.

- Isaacson, R. L., McKeachie, W. J., & Milholland, J. E. (1963). Correlation of teacher personality variables and student ratings. *Journal of Educational Psychology*, 54, 110–117. <http://doi.org/10.1037/h0048797>
- Jacob, R., Zhu, P., & Bloom, H. (2010). New empirical evidence for the design of group randomized trials in education. *Journal of Research on Educational Effectiveness*, 3, 157–198. <http://doi.org/10.1080/19345741003592428>
- John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The big five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, 61, 521–551. <http://doi.org/10.1111/j.1467-6494.1993.tb00781.x>
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–139). New York, NY: Guilford Press.
- Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, 98, 875–925. <http://doi.org/10.1037/a0033901>
- Kammeyer-Mueller, J. D., Livingston, B. A., & Liao, H. (2011). Perceived similarity, proactive adjustment, and organizational socialization. *Journal of Vocational Behavior*, 78, 225–236. <https://doi.org/10.1016/j.jvb.2010.09.012>
- Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, 27, 615–631. <http://doi.org/10.1016/j.econedurev.2007.05.005>
- Klassen, R. M., & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching

effectiveness: A meta-analysis. *Educational Research Review*, 12, 59–76.

<http://doi.org/10.1016/j.edurev.2014.06.001>

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality*, 64, 311–337.

<http://doi.org/10.1111/j.1467-6494.1996.tb00513.x>

Lala, V., & Priluck, R. (2011). When students complain: An antecedent model of students' intention to complain. *Journal of Marketing Education*, 33, 236–252.

<http://doi.org/10.1177/0273475311420229>

Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology*, 93, 268–279. <http://doi.org/10.1037/0021-9010.93.2.268>

MacCann, C., Joseph, D. L., Newman, D. A., & Roberts, R. D. (2014). Emotional intelligence is a second-stratum factor of intelligence: Evidence from hierarchical and bifactor models. *Emotion*, 14, 358–374. <http://doi.org/10.1037/a0034755>

McAbee, S. T., & Connelly, B. S. (2016). A multi-rater framework for studying personality: The Trait-Reputation-Identity Model. *Psychological Review*, 123, 569–591. <https://doi.org/10.1037/rev0000035>

Marsh, H. W. (2007). Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness. In R. P. Perry & J. C. Smart (Eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective* (pp. 319–383). Springer Netherlands.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707–754. <http://dx.doi.org/10.1037/0022-0663.76.5.707>

Marsh, H. W. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students'

- evaluations of university teaching. *British Journal of Educational Psychology*, 52, 77–95. <http://doi.org/10.1111/j.2044-8279.1982.tb02505.x>
- Marsh, H. W., & Overall, J. U. (1980). Validity of students' evaluations of teaching effectiveness: Cognitive and affective criteria. *Journal of Educational Psychology*, 72, 468–475. <http://dx.doi.org/10.1037/0022-0663.72.4.468>
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52, 1187–1197. <http://doi.org/10.1037/0003-066x.52.11.1187>
- Murray, H. G. (1975). Predicting student ratings of college teaching from peer ratings of personality types. *Teaching of Psychology*, 2(2), 66–69. http://doi.org/10.1207/s15328023top0202_4
- Murray, H. G., Rushton, J. P., & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology*, 82, 250–261. <http://doi.org/10.1037/0022-0663.82.2.250>
- OECD (2013). *PISA 2012 Results: Ready to Learn: Students' Engagement, Drive and Self-Beliefs (Volume III)*. Paris: OECD Publishing.
- Patrick, C. L. (2011). Student evaluations of teaching: Effects of the Big Five personality traits, grades and the validity hypothesis. *Assessment and Evaluation in Higher Education*, 36, 239–249. <http://doi.org/10.1080/02602930903308258>
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). New York: Guilford.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903. <http://doi.org/10.1037/0021->

9010.88.5.879

- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135, 322–338. <http://doi.org/10.1037/a0014996>
- Poropat, A. E. (2014). Other-rated personality and academic performance: Evidence and implications. *Learning and Individual Differences*, 34, 24–32.
<http://doi.org/10.1016/j.lindif.2014.05.013>
- Richardson, J. T. E. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education*, 30, 387–415.
<http://doi.org/10.1080/02602930500099193>
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138, 353–387. <http://doi.org/10.1037/a0026838>
- Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, 6, 43–74.
http://doi.org/10.1162/EDFP_a_00022
- Sabbe, E., & Aelterman, A. (2007). Gender in teaching: A literature review. *Teachers and Teaching*, 13, 521–538. <http://doi.org/10.1080/13540600701561729>
- Salgado, J. E. (1997). The Five Factor Model of personality and job performance in the European community, 82, 30–43. <http://doi.org/10.1037/0021-9010.82.1.30>
- Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment*, 63, 506–516.
http://doi.org/10.1207/s15327752jpa6303_8
- Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating Individual Students' Perceptions of Instructional Quality: An Investigation of their Factor Structure, Measurement Invariance, and Relations to Educational Outcomes. *Frontiers in Psychology*,

7(February), 1-16. <https://doi.org/10.3389/fpsyg.2016.00110>

Simpson, P. M., & Siguaw, J. A. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education*, 22, 199-213.

<http://doi.org/10.1177/0273475300223004>

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87, 245–251. <http://doi.org/10.1037/0033-2909.87.2.245>

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance.

American Economic Review, 102, 3628–3651. <http://doi.org/10.1257/aer.102.7.3628>

UNESCO Institute for Statistics. (2016). Global flow of tertiary-level students. Retrieved

March 4, 2016, from <http://www.uis.unesco.org/Education/Pages/international-student-flow-viz.aspx>

Table 1

Reliability and Descriptive Statistics for Student Personality (N = 515), Mean Student-Reported Instructor Personality (n = 88), and Self-Reported Instructor Personality (n = 45) with Comparisons between the Two Sources of Instructor Personality Report

	Student Personality			Instructor Personality								
	Self-Report			Mean Student-Report				Self-Report			Mean Student-Report vs. Self-Report	
	α	M	SD	α	M	SD	ICC	α	M	SD	g	r
Openness	0.78	48.50	8.64	0.81	47.69	3.61	0.08	0.79	48.53	7.79	-0.15	.01
Conscientiousness	0.85	43.54	10.54	0.88	53.08	4.28	0.10	0.86	46.80	9.64	0.55**	.32*
Extraversion	0.83	38.63	10.71	0.83	43.67	5.71	0.21	0.87	40.18	11.03	0.09	-.02
Agreeableness	0.84	51.56	8.72	0.89	52.37	5.32	0.18	0.88	51.13	9.12	-0.02	.19
Emotional Stability	0.82	36.85	10.58	0.79	46.87	3.89	0.02	0.90	38.89	11.63	0.69**	.08

Note. r = r effect size, g = Hedge's g , whose significance was determined by paired samples t -test.

ICC = A measure of inter-rater reliability using Spearman-Brown adjustments of the intraclass correlation.

* $p < .05$.

Table 2

Reliability and Descriptive Statistics of Student Educational Outcomes and their Correlations with Student Personality (N = 515), Mean Student-Reported Instructor Personality (n= 88), and Self-Reported Instructor Personality (n = 45)

	α	M	SD	Student Personality					Instructor Personality									
				Self-Report					Mean Student-Report					Self-Report				
				O	C	E	A	ES	O	C	E	A	ES	O	C	E	A	ES
SEEQ																		
Learning Focus	0.79	15.41	3.13	.16**	.12**	.08	.09*	.05	.34**	.14	.29**	.17	.16	-.10	-.14	-.06	-.15	-.17
Enthusiasm	0.90	14.26	3.65	.12**	.08	.13**	.15**	.16**	.36**	.57**	.66**	.64**	.52**	.08	-.08	-.06	-.13	-.20
Organization	0.77	8.14	1.59	.11*	.06	.01	.17**	.07	.32**	.51**	.41**	.34**	.37**	.05	-.17	-.18	-.07	-.22
Group Interaction	0.90	15.51	3.55	-.04	.03	.07	.15**	.14**	.31**	.41**	.50**	.48**	.49**	.01	-.03	-.06	.04	-.07
Individual Rapport	0.86	15.92	3.03	.08	.14**	.05	.23**	.15**	.37**	.57**	.46**	.68**	.46**	.00	.10	-.12	.03	-.18
Overall Course Evaluation	-	3.74	0.95	.12**	.06	.02	.12**	.03	.30*	.04	.35**	.26*	.17	-.06	-.14	.01	-.14	-.28
Overall Instructor Evaluation	-	3.94	0.93	.08	.05	.04	.13**	.07	.31**	.61**	.53**	.59**	.54**	-.04	.15	-.08	.02	-.20
PSE		69.83	9.49	.26**	.25**	.06	.02	.05	.31**	.05	.15	-.01	.06	-.11	-.14	-.31*	-.18	-.33*
Course Mark ^a	-	63.90	15.37	.05	.20**	-.04	-.05	.01	.02	-.09	-.02	-.07	.08	-.15	-.21	-.14	-.06	-.03

Note. O = Openness, C = Conscientiousness, E = Extraversion, A = Agreeableness, ES = Emotional Stability, SEEQ = Student Evaluation of Educational Quality, PSE = Performance Self-Efficacy.

^a n = 456 for student personality.

* $p < .05$, ** $p < .01$.

Table 3

Correlations between Student-Reported Instructor Personality and Student Educational Outcomes using Random Assignment Design (n = 2009) and Single Student-Report Design (n = 88)

	Random Assignment Design ^a					Single Student-Report Design ^b				
	O	C	E	A	ES	O	C	E	A	ES
SEEQ										
Learning Focus	.26**	.26**	.20**	.22**	.21**	.31**	.14	.22	.20	.16
Enthusiasm	.46**	.40**	.54**	.48**	.26**	.34**	.46**	.52**	.59**	.39**
Organization	.38**	.49**	.33**	.35**	.25**	.34	.40**	.32**	.33**	.28*
Group Interaction	.32**	.37**	.36**	.40**	.27**	.22	.24*	.37**	.40**	.29*
Individual Rapport	.37**	.37**	.32**	.60**	.33**	.32**	.42**	.39**	.58**	.24*
Overall Course Evaluation	.25**	.28**	.29**	.26**	.20**	.25*	.04	.25*	.29*	.11
Overall Instructor Evaluation	.38**	.43**	.42**	.48**	.28**	.27*	.47**	.32**	.49**	.36**
PSE	.16**	.10*	.13**	.10*	.06	.21	.05	.04	.06	.10
Course Mark	n/a	n/a	n/a	n/a	n/a	.03	-.03	.01	.03	.12

Note. O = Openness, C = Conscientiousness, E = Extraversion, A = Agreeableness, ES = Emotional Stability, SEEQ = Student Evaluation of Educational Quality, PSE = Performance Self-Efficacy.

^a In the random assignment design, outcome data for each instructor was randomly assigned amongst the students within an instructor (i.e., personality and outcome data did not come from the same students).

^b In the single student-report design, a single student was randomly selected for each instructor, and their ratings of the instructor personality were used (i.e., student-reports of instructor personality were obtained from a single rater, so as to be comparable to instructor self-reports in terms of the number of raters).

* $p < .05$. ** $p < .01$.

Table 4

Multilevel Multiple Regression Analyses Predicting Student Educational Outcomes from the Covariates, Self-Reported Student Personality, and Mean Student-Reported Instructor Personality in Standardized Regression Coefficients with R-Squares in Parentheses (n = 299)

	SEEQ							PSE	Course Mark ^a
	Learning Focus	Enthusiasm	Organization	Group Interaction	Individual Rapport	Overall Course Evaluation	Overall Instructor Evaluation		
Level 1 (Student)	(.09 ^{**})	(.05 [*])	(.08 ^{**})	(.03)	(.05)	(.08 ^{**})	(.04 [*])	(.17 ^{**})	(.17 ^{**})
Gender	0.08	0.01	0.04	0.00	-0.01	0.10	0.06	-0.06	0.04
Age	0.14 [*]	-0.03	-0.04	-0.03	-0.09	0.12 [*]	0.02	-0.03	-0.03
CA	0.14 [*]	0.07	0.20 ^{**}	0.03	0.06	0.11	0.08	0.14 [*]	0.29 ^{**}
O	0.14 ^{**}	0.09	0.07	-0.06	0.07	0.12	0.08	0.19 ^{**}	-0.01
C	0.03	0.04	-0.03	-0.03	0.11	-0.01	0.00	0.29 ^{**}	0.31 ^{**}
E	0.08	0.11	-0.07	-0.02	-0.01	0.02	-0.02	0.04	-0.07
A	0.06	0.04	0.18 ^{**}	0.16 ^{**}	0.09	0.12	0.12 [*]	-0.10	-0.13
ES	0.04	0.04	-0.02	.03	-0.02	-0.06	0.02	0.00	0.00
Level 2 (Instructor)	(.97 ^{**})	(1.00)	(.91 ^{**})	(.91 ^{**})	(.99 ^{**})	(.99 ^{**})	(.99 ^{**})	(1.00 ^{**})	(.45)
Gender	0.60 ^{**}	-0.19	-0.26	0.21	0.12	0.53 ^{**}	0.03	0.29	0.25
Age	0.09	-0.05	-0.03	-0.16	0.32 ^{**}	-0.01	-0.02	0.00	0.09
O	0.59	0.04	0.21	0.40 [*]	-0.03	0.81 ^{**}	0.01	0.87	0.71 [*]
C	0.22	0.22	0.64 ^{**}	-0.02	0.33	-0.20	0.51 ^{**}	-0.40	-0.31
E	-0.23	0.76	0.41 [*]	0.17	0.31 [*]	0.02	0.33 [*]	-0.49	-0.47
A	-0.15	0.60 ^{**}	0.45 [*]	0.00	0.97 ^{**}	-0.10	0.83 ^{**}	0.58	-0.30
ES	0.34	-0.27	-0.42	0.53	-0.44 [*]	0.28	-0.43	-0.58	0.24

Note. R^2 are in parentheses.

CA = Cognitive Ability, SEEQ = Student Evaluation of Educational Quality, PSE = Performance Self-Efficacy, O = Openness, C = Conscientiousness, E = Extraversion, A = Agreeableness, ES = Emotional Stability.

* $p < .05$. ** $p < .01$.

Table 5

Multilevel Multiple Regression Analyses Predicting Student Educational Outcomes from the Covariates, Self-Reported Student Personality, and Self-Reported Instructor Personality in Standardized Regression Coefficients with R-Squares in Parentheses (n = 299)

	SEEQ							PSE	Course Mark ^a
	Learning Focus	Enthusiasm	Organization	Group Interaction	Individual Rapport	Overall Course Evaluation	Overall Instructor Evaluation		
Level 1 (Student)	(.10 ^{**})	(.07 ^{**})	(.10 ^{**})	(.04)	(.07)	(.08)	(.05 ^{**})	(.18 ^{**})	(.16 ^{**})
Gender	0.05	0.01	0.04	-0.03	-0.04	0.08	0.04	-0.08	0.02
Age	0.11 [*]	-0.05	-0.06	-0.06	-0.10	0.09	0.01	-0.03	-0.03
CA	0.13 [*]	0.05	0.20 ^{**}	0.02	0.06	0.10	0.06	0.14	0.28 ^{**}
O	0.14 ^{**}	0.13 [*]	0.10 [*]	-0.05	0.09 [*]	0.12 [*]	0.11	0.19 ^{**}	-0.01
C	0.03	0.03	-0.01	-0.01	0.12	-0.01	0.03	0.29 ^{**}	0.31 ^{**}
E	0.11	0.11	-0.06	0.00	0.00	0.05	-0.02	0.06	-0.05
A	0.08	0.06	0.18 ^{**}	0.18 [*]	0.09	0.14	0.11	-0.08	-0.11
ES	0.03	0.05	0.00	0.04	-0.01	-0.06	0.05	-0.01	-0.01
Level 2 (Instructor)	(.88)	(.34)	(.15)	(.46)	(.74 ^{**})	(.96)	(.30)	(.90)	(.30)
Gender	1.02	0.31	0.05	0.55	0.57 ^{**}	0.90	0.28	0.13	0.01
Age	0.21	0.14	0.05	-0.24	0.66 ^{**}	0.09	-0.05	0.16	0.12
O	0.09	0.33 [*]	0.27	0.17	0.32	0.23	0.21	-0.06	-0.27
C	-0.26	-0.30	-0.11	0.07	-0.07	-0.21	0.14	-0.72	-0.41
E	-0.05	0.03	-0.27	-0.37	-0.02	0.09	-0.09	-0.27	0.23
A	-0.34	-0.10	0.09	0.25	0.21	-0.12	0.26	0.53	-0.02
ES	0.12	-0.24	-0.17	-0.07	-0.53	-0.33	-0.41	-0.57	0.29

Note. R^2 are in parentheses.

CA= Cognitive Ability, SEEQ = Student Evaluation of Educational Quality, PSE = Performance Self-Efficacy, O = Openness, C = Conscientiousness, E = Extraversion, A = Agreeableness, ES = Emotional Stability. ^an = 267.

* $p < .05$, ** $p < .01$.